

**METHOD OF DESIGNING AGONISTS AND ANTAGONISTS TO  
EGF RECEPTOR FAMILY**

**Field of the Invention**

This invention relates to the field of epidermal growth factor (EGF) receptor structure and EGF receptor/ligand interactions. In particular, it relates to the field of using the EGF receptor structure to select and screen for ligands of the EGF receptor.

**Background of the Invention**

Epidermal growth factor is a small polypeptide cytokine that stimulates marked proliferation of epithelial tissues and is a member of a larger family of structurally related cytokines such as transforming growth factor  $\alpha$  (TGF $\alpha$ ), amphiregulin, betacellulin, heparin-binding EGF and some viral gene products. Abnormal EGF family signalling is a characteristic of certain cancers (Soler, C. & Carpenter, G., 1994 In Nicola, N. (ed) "Guidebook to Cytokines and their Receptors", Oxford Univ. Press, Oxford, pp194-197; Walker, F. & Burgess, A. W., 1994, In Nicola, N. (ed) "Guidebook to Cytokines and their Receptors", Oxford Univ. Press, Oxford, pp198-201).

The epidermal growth factor receptor (EGFR) is the cell membrane receptor for EGF (Ullrich, A., and Schlessinger, J. (1990) *Cell* 61, 203-212). The EGFR also binds other ligands that contain amino acid sequences classified as the EGF-like motif. Among these ligands, the three-dimensional structures of EGF and TGF $\alpha$  have been determined by NMR (Montelione, G.T.; Wuthrich, K.; Nice, E.C., Burgess, A.W. and Scheraga, H.A. (1986) *PNAS* 83(22): 8594-8; Campbell, I.D., Cooke, R.M., Baron, M., Harvey, T.S., and Tappin, M.J. (1989) *Prog. Growth Factor Res.* 1, 13-22). Upon binding of the ligand to the extracellular domain, the EGFR undergoes dimerization, which eventually leads to the activation of its cytoplasmic protein tyrosine kinase (Ullrich, A., and Schlessinger, J. (1990) *Cell* 61, 203-212). The EGFR is also known as the ErbB-1 receptor and belongs to the type I family of receptor tyrosine kinases (Ullrich, A., and Schlessinger, J. (1990) *Cell* 61, 203-212). This group also includes the ErbB-2, ErbB-3 and ErbB-4 receptors. The ligand of ErbB-2 is still unknown but it is clear that heregulin binds to ErbB-3 and ErbB-4 (Plowman, G.D., Green, J.M., Calouscou, J.M., Carlton, G.W., Rothwell, V.M., and Buckley, S. (1993) *Nature* 366, 473-475). One of the heregulins is known as neuregulin or NDF and contains an EGF-like sequence that was found to fold into an EGF-like fold by NMR (Nagata, K.,

Kohda, D., Hatanska, H., Ichikawa, S., Matsuda, S., Yamamoto, T., Suzuki, A., and Inagaki, F. (1994) *EMBO J.* 13, 3517-3523 and Jacobson, N.E., Abadl, N., Sliwowski, M.X., Reilly, D., Skelton, N.J., and Fairbrother, W.J. (1996) *Biochemistry* 36, 3402-3417).

5       The type II family of receptor tyrosine kinases consists of the insulin receptor (INSR), the insulin-like growth factor I receptor (IGF-1), and the insulin receptor-related receptor (Ullrich, A., and Schlessinger, J. (1990) *Cell* 61, 203-212). Although the type II receptors consist of four chains ( $\alpha_2\beta_2$ ), both the extracellular portions of the receptors from the two families, as well  
10 as the tyrosine kinase portions, share significant sequence homology, suggesting a common evolutionary origin (Ullrich, A., and Schlessinger, J. (1990) *Cell* 61, 203-212, and Bajaj, M., Waterfield, M.D., Schlessinger, J., Taylor, W.R., and Blundell, T. (1987) *Biochim. Biophys. Acta* 916, 220-226).

15       The 621 amino acid residues of the extracellular domain of the human EGFR (sEGFR) can be subdivided into four domains as follows: L1, S1, L2 and S2, where L and S stand for "large" and "small" domains, respectively (Bajaj, M., Waterfield, M.D., Schlessinger, J., Taylor, W.R., and Blundell, T. (1987) *Biochim. Biophys. Acta* 916, 220-226, see Fig. 2). The L1 and L2 domains are homologous, as are the S1 and S2 domains.

20       Ligand-induced dimerization was first reported for the EGF receptor (Schlessinger, J. (1980) *Trends Biochem Sci* 13, 443-447) and now is widely accepted as a general mechanism for the transmission of growth stimulatory signals across the cell membrane. Although many biochemical experiments have been performed to reveal the molecular mechanism of receptor  
25 dimerization (Lemmon, M.A., Bu, Z., Ladbury, J.E., Zhou, M., Pinchasi, D., Lax, L., Engelman, D.M., and Schlessinger, J. (1997) *EMBO J.* 16, 281-294 and Tzabar, E., Pinkas-Kramarski, R., Moyer, J.D., Klapper, D.N., Alroy, L., Levkowitz, G., Shelly, M., Henis, S., Eisenstein, M., Ratzkin, B.J., Sela, M., Andrews, G.C., and Yarden, Y. (1997) *EMBO J.* 16, 4938-4950 and Lax, L.,  
30 Mitra, A.K., Ravern, C., Hurwitz, D.R., Rubinstein, M., Ullrich, A., Stroud, R.M., and Schlessinger, J. (1991), *J. Biol. Chem.* 266, 13828-13833), the molecular mechanism by which monomeric ligands induce dimerization is still unknown for members of the EGFR family. Single particle averaging of electron microscopic images suggests that the overall shape of the sEGFR is  
35 four-lobed and doughnut-like (Lax, L., Mitra, A.K., Ravern, C., Hurwitz, D.R., Rubinstein, M., Ullrich, A., Stroud, R.M., and Schlessinger, J. (1991), *J. Biol.*

*Chem.* 266, 13828-13833). Small angle x-ray scattering also indicates that the sEGFR is a flattened sphere with long diameters of 110 Å and a short diameter of 20 Å (Lemmon, M.A., Bu, Z., Ladbury, J.E., Zhou, M., Pinchasi, D., Lax, L., Engelman, D.M., and Schlessinger, J. (1997) *EMBO J.* 16, 281-294).

5 The crystallization of sEGFR in complex with EGF has been published (Günther, N., Betzel, C., and Weber, W. (1990) *J. Biol. Chem.* 265, 22082-22085; Degenhardt M., Weber W., Eschenburg S., Dierks K., Funari SS., Rapp G. and Betzel C. (1998) *Acta Crystallogr. D Biol. Crystallogr.* 54:999-1001), but the structure has not yet been reported, despite a decade of effort by many  
10 groups.

One EGF receptor ligand, TGF- $\alpha$  has been observed to be overproduced in keratinocyte cells which are subject to psoriasis (Turbitt, M.L. et al., 1990, *J. Invest. Dermatol.* 95(2), 229-232; Higashimiyama, M. et al., 1991, *J. Dermatol.*, 18(2), 117-119; Elder, J.T. et al, 1990, 94(1), 19-25). The  
15 overproduction of at least one other EGF receptor ligand, amphiregulin, has also been implicated in psoriasis. (Piepkorn, M. 1996, *Am. J. Dermatopath.*, 18(2), 165-171). Molecules that inhibit the EGF receptor have been shown to inhibit the proliferation of both normal keratinocytes (Dvir, A. et al, 1991, *J. Cell Biol.*, 113(4), 857-865) and psoriatic keratinocytes. (Ben-Bassat, H. et al.,  
20 1995, *Exp. Dermatol.*, 4(2), 82-88). These findings indicate that EGF receptor antagonists may be useful in the treatment of psoriasis.

Many cancer cells express constitutively active EGFR (Sandgreen, E. P., et al., 1990, *Cell*, 61:1121-135; Karnes, W. E. J., et al., 1992, *Gastroenterology*, 102:474-485) or other EGFR family members (Hynes, N. E., 1993, *Semin. Cancer Biol.* 4:19-26). Elevated levels of activated EGFR  
25 occur in bladder, breast, lung and brain tumours (Harris, A. L., et al., 1989, In Furth & Greaves (eds) *The Molecular Diagnostics of human cancer*. Cold Spring Harbor Lab. Press, CSH, NY, pp353-357). Antibodies to EGFR can inhibit ligand activation of EGFR (Sato, J. D., et al., 1983 *Mol. Biol. Med.*  
30 1:511-529) and the growth of many epithelial cell lines (Aboud-Pirak E., et al., 1988, *J. Natl Cancer Inst.* 85:1327-1331). Patients receiving repeated doses of a humanised chimeric anti-EGFR monoclonal antibody (Mab) showed signs of disease stabilization. The large doses required and the cost of production of humanised Mab is likely to limit the application of this type of  
35 therapy. These findings indicate that the development of EGF receptor antagonists will be attractive anticancer agents.

### Summary of the Invention

The present inventors have now obtained three-dimensional structural information concerning the epidermal growth factor receptor (EGFR). This structural information was obtained by comparative modelling based on the three-dimensional structure of the IGF-1 receptor as described in PCT/AU98/00998. The information presented in the present application can be used to predict the structure of related members of the EGF receptor family, and to develop specific ligands of members of the EGF receptor family for therapeutic applications.

Accordingly, in a first aspect the present invention provides a method of designing a compound which binds to a molecule of the EGF receptor family and modulates an activity mediated by the molecule, which method comprises the step of assessing the stereochemical complementarity between the compound and a topographic region of the molecule, wherein the molecule is characterised by

(i) amino acids 1-621 of the EGF receptor positioned at atomic coordinates substantially as shown in Figure 6;

(ii) one or more subsets of said amino acids related to the coordinates shown in Figure 6 by whole body translations and/or rotations; or

(iii) amino acids present in the amino acid sequence of a member of the EGF receptor family, which form an equivalent three-dimensional structure to that of the receptor site defined by amino acids 1-621 of the EGF receptor positioned at atomic coordinates substantially as shown in Figure 6.

In a preferred embodiment of the first aspect, the topographic region of the molecule is defined by amino acids 1-475 of the EGF receptor, or an amino acid sequence which forms an equivalent three-dimensional structure to that of the region defined by amino acids 1-475 of the EGF receptor positioned at atomic coordinates substantially as shown in Figure 6.

In a further preferred embodiment of the first aspect, the topographic region of the molecule is defined by amino acids 313-621 of the EGF receptor, or an amino acid sequence which forms an equivalent three-dimensional structure to that of the region defined by amino acids 313-621 of the EGF receptor positioned at atomic coordinates substantially as shown in Figure 6.

The phrase "EGF receptor family" includes, but is not limited to, the EGF receptor, ErbB2, ErbB3 and ErbB4. In general, EGF receptor family

molecules show similar domain arrangements and share significant sequence identity, preferably at least 40% identity.

The EGF receptor molecule defined in the first aspect of the present invention is depicted in Figure 5. The fragment comprising residues 1-475 of the receptor comprises the L1, S1 and L2 domains of the ectodomain of the EGF receptor. At the centre of this structure is a cavity, bounded by all three domains, of sufficient size to accommodate a ligand molecule.

The fragment comprising residues 313-621 comprises the L2 and S2 domains, which are positioned such that they form a "corner" structure. It is envisaged that this corner structure provides a further binding site for ligands of EGF receptor family members.

By "stereochemical complementarity" we mean that the substance or a portion thereof correlates, in the manner of the classic "lock-and-key" visualisation of ligand-receptor interaction, with the cavity in the receptor site.

In a preferred embodiment of the first aspect of the present invention, the method further involves selecting or designing a compound which has portions that match residues positioned on the surface of the receptor site as depicted in Figures 7, 8 and 9. By "match" we mean that the identified portions interact with the surface residues, for example, via hydrogen bonding or by enthalpy-reducing Van der Waals interactions which promote desolvation of the biologically active compound within the site, in such a way that retention of the compound within the cavity is favoured energetically.

In a further preferred embodiment of the first aspect of the present invention, the method includes screening for, or designing, a compound which possesses a stereochemistry and/or geometry which allows it to interact with both the L1 and L2 domains of the receptor site. It is believed that EGFR monomers may dimerise in nature in such a manner that the cavities of each monomer may face each other. Accordingly, the method of the first aspect of the present invention may involve screening for, or designing, a biologically active compound which interacts with the L1 domain of one monomer and the L2 domain of the other monomer.

In a further preferred embodiment of the first aspect of the present invention the compound interacts with a fragment in the region of the L1 domain-S1 domain interface, causing an alteration in the positions of the

domains relative to each other. Preferably, the interaction of the compound causes the L1 and S1 domains to move away from each other. In a further preferred embodiment the compound interacts with the hinge region between the S1 domain and the L2 domain causing an alteration in the positions of these domains relative to each other. In a further preferred embodiment the compound interacts with the  $\beta$  sheet of the L1 domain causing an alteration in the position of the L1 domain relative to the position of the S1 domain or L2 domain.

In a further preferred embodiment, the compound binds to a lower face (according to orientations shown in Figures 3 and 4) containing the second  $\beta$ -sheet of the L1 and/or L2 domains, wherein the structure of the face is characterised by a plurality of solvent-exposed hydrophobic residues. Examples of these hydrophobic residues include Tyr64, Leu66, Tyr89, Tyr93 (see Figure 7), Leu348, Phe380 and Phe412 (see Figure 10).

In a further preferred embodiment the compound interacts with the hinge region between the L2 domain and S2 domains, causing an alteration in the positions of the L1 and L2 domains relative to each other. Preferably, the interaction of the compound causes the L1 and L2 domains to move away from each other.

In a further preferred embodiment the compound interacts with the  $\beta$  sheet of the L2 domain causing an alteration in the position of the L2 domain relative to the position of the L1 domain.

In a further preferred embodiment of the present invention, the stereochemical complementarity is such that the compound has a  $K_d$  for the receptor site of less than  $10^{-6}$ M. More preferably, the  $K_d$  value is less than  $10^{-8}$ M and more preferably less than  $10^{-9}$ M.

In preferred embodiments of the first aspect of the present invention, the compound is selected or modified from a known compound identified from a data base.

In one embodiment of the first aspect, the compound has the ability to increase an activity mediated by the molecule of the EGF receptor family.

In another embodiment, the compound has the ability to decrease an activity mediated by the molecule of the EGF receptor family. Preferably, the stereochemical interaction between the compound and the receptor site is adapted to prevent the binding of a natural ligand of the molecule of the EGF

receptor family to the receptor site. Preferably, the compound has a  $K_i$  of less than  $10^{-6}$ M, more preferably less than  $10^{-8}$ M and more preferably less than  $10^{-9}$ M.

In a second aspect the present invention provides computer-assisted method for identifying potential compounds able to bind to a molecule of the EGF receptor family and to modulate an activity mediated by the molecule, using a programmed computer comprising a processor, an input device, and an output device, comprising the steps of:

(a) inputting into the programmed computer, through the input device, data comprising the atomic coordinates of the EGF receptor molecule as shown in Figure 6, or a subset thereof;

(b) generating, using computer methods, a set of atomic coordinates of a structure that possesses stereochemical complementarity to the atomic coordinates of the EGF receptor site as shown in Figure 6, or a subset thereof, thereby generating a criteria data set;

(c) comparing, using the processor, the criteria data set to a computer database of chemical structures;

(d) selecting from the database, using computer methods, chemical structures which are similar to a portion of said criteria data set; and

(e) outputting, to the output device, the selected chemical structures which are similar to a portion of the criteria data set.

In a preferred embodiment of the second aspect, the method is used to identify potential compounds which have the ability to decrease an activity mediated by the receptor.

In a further preferred embodiment of the second aspect, the method further comprises the step of selecting one or more chemical structures from step (e) which interact with the receptor site of the molecule in a manner which prevents the binding of natural ligands to the receptor site.

In a further preferred embodiment of the second aspect, the method further comprises the step of obtaining a compound with a chemical structure selected in steps (d) and (e), and testing the compound for the ability to decrease an activity mediated by the receptor.

In a further preferred embodiment of the second aspect, the method is used to identify potential compounds which have the ability to increase an activity mediated by the receptor molecule.

In a further preferred embodiment of the second aspect, the method further comprises the step of obtaining a molecule with a chemical structure

selected in steps (d) and (e), and testing the compound for the ability to increase an activity mediated by the receptor molecule.

The present invention also provides a method of screening of a putative compound having the ability to modulate the activity of a molecule of the EGF receptor family, comprising the steps of identifying a putative compound by a method according to the first or second aspects, and testing the compound for the ability to increase or decrease an activity mediated by the molecule. In one embodiment, the test is carried out *in vitro*. Preferably, the *in vitro* test is a high throughput assay. In another embodiment, the test is carried out *in vivo*.

In a third aspect the present invention provides a compound able to bind to a molecule of the EGF receptor family and to modulate an activity mediated by the molecule, the compound being obtained by a method according to the present invention.

In a preferred embodiment of the third aspect, the compound is a mutant ligand of a molecule of the EGF receptor family, where at least one mutation occurs in the region of the ligand which interacts with residues on the surface of the receptor site facing toward the cavity. For example, the residues Arg 41 and Tyr 13 in EGF are conserved in other members of the EGF receptor family of ligands (a Phe residue may be substituted for Tyr 13). Structures of several EGF family members show the two residues to be in close proximity (Groenen, L.C., Nice, E.C., Burgess, A.W., 1994, Growth Factors 11:235-257). This portion of EGF may interact with a hydrophobic portion of the EGF receptor which contains one or more negatively charged residues such as the lower  $\beta$  sheet of the L1 domain. Mutants of EGF which show altered activity may be generated by introducing modifications to Arg 41 or Tyr 13 or other nearby residues. Alternatively, mutants of EGF may be generated by introducing modifications to residues on the opposite side of the ligand which may interact with a second receptor molecule in the unmodified ligand.

In a fourth aspect the present invention provides a compound which possesses stereochemical complementarity to a topographic region of a molecule of the EGF receptor family and modulates an activity mediated by the molecule, wherein the molecule is characterised by

(i) amino acids 1-621 of the EGF receptor positioned at atomic coordinates substantially as shown in Figure 6;



(ii) one or more subsets of said amino acids related to the coordinates shown in Figure 6 by whole body translations and/or rotations; or

(iii) amino acids present in the amino acid sequence of a member of the EGF receptor family, which form an equivalent three-dimensional structure to that of the receptor site defined by amino acids 1-621 of the EGF receptor positioned at atomic coordinates substantially as shown in Figure 6;

with the proviso that the compound is not a naturally occurring ligand of a molecule of the EGF receptor family or a mutant thereof.

By "mutant" we mean a ligand which has been modified by one or more point mutations, insertions of amino acids or deletions of amino acids.

In a preferred embodiment of the fourth aspect, the topographic region of the molecule is defined by amino acids 1-475 of the EGF receptor or an amino acid sequence which forms an equivalent three-dimensional structure to that of the region defined by amino acids 1-475 of the EGF receptor positioned at atomic coordinates substantially as shown in Figure 6.

In a further preferred embodiment of the fourth aspect, the topographic region of the molecule is defined by amino acids 313-621 of the EGF receptor or an amino acid sequence which forms an equivalent three-dimensional structure to that of the region defined by amino acids 313-621 of the EGF receptor positioned at atomic coordinates substantially as shown in Figure 6.

In preferred embodiments of the third and fourth aspects, the stereochemical complementarity between the compound and the receptor site is such that the compound has a  $K_d$  for the receptor site of less than  $10^{-6}M$ , more preferably less than  $10^{-8}M$ .

In some embodiments of the third and fourth aspects, the compound increases an activity mediated by the EGF receptor.

In other embodiments of the third and fourth aspects, the compound decreases an activity mediated by the EGF receptor.

In a fifth aspect, the present invention provides a pharmaceutical composition for preventing or treating a disease which would benefit from increased signalling by a molecule of the EGF receptor family, which comprises a compound according to the third or fourth aspects of the present invention and a pharmaceutically acceptable carrier or diluent.

In a sixth aspect, the present invention provides a pharmaceutical composition for preventing or treating a disease associated with signalling by a molecule of the EGF receptor family which comprises a compound

according to the third or fourth aspects of the present invention and a pharmaceutically acceptable carrier or diluent.

5 In a seventh aspect the present invention provides a method of preventing or treating a disease which would benefit from increased signalling by a molecule of the EGF receptor family which method comprises administering to a subject in need thereof a compound according to the third or fourth aspects of the present invention. Preferably, the disease is selected from wound healing and gastric ulcers.

10 In an eighth aspect the present invention provides a method of preventing or treating a disease associated with signalling by a molecule of the EGF receptor family which method comprises administering to a subject in need thereof a compound according to the third or fourth aspects of the present invention. Preferably, the disease is selected from psoriasis and tumour states comprising but not restricted to cancer of the breast, brain,  
15 ovary, cervix, pancreas, lung, head and neck, and melanoma, rhabdomyosarcoma, mesothelioma and glioblastoma.

Throughout this specification, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but  
20 not the exclusion of any other element, integer or step, or group of elements, integers or steps.

### **Brief Description of the Drawings**

25 **Figure 1:** Sequence alignment of human EGF receptor family proteins with IGF-1 receptor sequences and insulin receptor sequence for the first two domains of the EGF receptor. The alignment of the EGF receptor and the various IGF-1 receptor sequences were used by the MODELLER program to create a model of the EGF receptor domains L1 and S1. Residues which are  
30 underlined were used to create additional  $\alpha$ - $\alpha$  restraints for the construction of the EGF receptor model. Disulfide bonds are also indicated by lines between cysteine residues. The modules of the EGF receptor S1 domain are numbered.

35 **Figure 2:** Sequence alignment of human EGF receptor family proteins with IGF-1 receptor sequences and insulin receptor sequence for the third and

fourth domains of the EGF receptor. Additional labels and lines are similar to those in figure 1.

5 **Figure 3:** Model polypeptide fold of the L1 and S1 domains of the EGF receptor. The L1 is at the left hand side of the structure with the N-terminus facing the front. Cysteine residue sidechains are depicted as sticks.

10 **Figure 4:** Model polypeptide fold of the L2 and S2 domains of the EGF receptor. The L2 is at the bottom of the structure with the N-terminus facing the front. Cysteine residue sidechains are depicted as sticks.

15 **Figure 5:** Superposition of the two models (of the L1 and S1 domain and of L2 and S2 domains) onto the structure of the first three domains of the IGF-1 receptor. Cysteine residue sidechains are depicted as sticks. Selected residues are shown as spheres and labelled.

SUB  
A1  
20 **Figure 6:** Coordinates of the two models of the EGF receptor extracellular domain. The first model consists of the domains L1 and S1. The second model consists of the domains L2 and S2. The coordinates are in relation to a Cartesian set of orthogonal axes. The L1, S1 and L2 domains of the EGF receptor models have been superimposed on the crystal structure of the IGF-1 receptor domains L1, cysteine-rich domain and L2. The final column contains the number 20, 40 or 60, depending on whether the residue containing the atom is judged to be well-modelled, have a moderate possibility of error, or is likely to be inaccurate, respectively.

30 **Figure 7:** Part of the model polypeptide fold of the L1 and S1 domains of the EGF receptor. Side chains of residues from the L1 domain which face towards the large cavity (shown in Figure 5) are shown in ball and stick notation and labelled with residue number and the one letter code.

35 **Figure 8:** Part of the model polypeptide fold of the L1 and S1 domains of the EGF receptor. Side chains of residues from the S1 domain which face towards the large cavity (shown in Figure 5) are shown in ball and stick notation and labelled using the one letter code.

Figure 9: Part of the model polypeptide fold of the L2 and S2 domains of the EGF receptor. Side chains of residues from the L2 domain which face towards the large cavity (shown in Figure 5) are shown in ball and stick notation and labelled using the one letter code.

5

Figure 10: Part of the model polypeptide fold of the L2 and S2 domains of the EGF receptor. Solvent exposed residues from the face of the L2 domain containing the large  $\beta$  sheet are shown in ball and stick representation.

## 10 Detailed description of Preferred Embodiments of the Invention

The present inventors have developed three dimensional structural information about the EGF receptor to enable a more accurate understanding of how the binding of ligand leads to signal transduction. Such information provides a rational basis for the development of ligands for specific therapeutic applications, something that heretofore could not have been predicted *de novo* from available sequence data.

The precise mechanisms underlying the binding of agonists and antagonists to the EGF receptor are not fully clarified. However, the binding of ligands to the receptor site, preferably with an affinity in the order of  $10^{-8}$ M or higher, is understood to arise from enhanced stereochemical complementarity relative to naturally occurring EGF receptor ligands.

Such stereochemical complementarity, pursuant to the present invention, is characteristic of a molecule that matches intra-site surface residues lining the groove of the receptor site as enumerated by the coordinates set out in Figure 6. The residues lining the groove are depicted in Figures 7, 8 and 9. By "match" we mean that the identified portions interact with the surface residues, for example, via hydrogen bonding or by enthalpy-reducing Van der Waals interactions which promote desolvation of the biologically active compound within the site, in such a way that retention of the biologically active compound within the groove is favoured energetically.

Substances which are complementary to the shape of the receptor site characterised by amino acids positioned at atomic coordinates set out in Figure 6 may be able to bind to the receptor site and, when the binding is sufficiently strong, substantially prohibit binding of the naturally occurring ligands to the site.

It will be appreciated that it is not necessary that the complementarity between ligands and the receptor site extend over all residues lining the groove in order to inhibit binding of the natural ligand. Accordingly, agonists or antagonists which bind to a portion of the residues lining the groove are encompassed by the present invention.

In general, the design of a molecule possessing stereochemical complementarity can be accomplished by means of techniques that optimize, either chemically or geometrically, the "fit" between a molecule and a target receptor. Known techniques of this sort are reviewed by Sheridan and Venkataraghavan, *Acc. Chem. Res.* 1987 20 322; Goodford, *J. Med. Chem.* 1984 27 557; Beddell, *Chem. Soc. Reviews* 1985, 279; Hol, *Angew. Chem.* 1986 25 767 and Verlinde C.L.M.J & Hol, *W.G.J. Structure* 1994, 2, 577, the respective contents of which are hereby incorporated by reference. See also Blundell et al., *Nature* 1987 326 347 (drug development based on information regarding receptor structure).

Thus, there are two preferred approaches to designing a molecule, according to the present invention, that complements the shape of the EGF receptor. By the geometric approach, the number of internal degrees of freedom (and the corresponding local minima in the molecular conformation space) is reduced by considering only the geometric (hard-sphere) interactions of two rigid bodies, where one body (the active site) contains "pockets" or "grooves" that form binding sites for the second body (the complementing molecule, as ligand). The second preferred approach entails an assessment of the interaction of respective chemical groups ("probes") with the active site at sample positions within and around the site, resulting in an array of energy values from which three-dimensional contour surfaces at selected energy levels can be generated.

The geometric approach is illustrated by Kuntz et al., *J. Mol. Biol.* 1982 161 269, the contents of which are hereby incorporated by reference, whose algorithm for ligand design is implemented in a commercial software package distributed by the Regents of the University of California and further described in a document, provided by the distributor, which is entitled "Overview of the DOCK Package, Version 1.0.", the contents of which are hereby incorporated by reference. Pursuant to the Kuntz algorithm, the shape of the cavity represented by the EGF receptor site is defined as a series of overlapping spheres of different radii. One or more extant databases of

crystallographic data, such as the Cambridge Structural Database System maintained by Cambridge University (University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.) and the Protein Data Bank maintained by Brookhaven National Laboratory (Chemistry Dept. Upton, NY 5 11973, U.S.A.), is then searched for molecules which approximate the shape thus defined.

Molecules identified in this way, on the basis of geometric parameters, can then be modified to satisfy criteria associated with chemical complementarity, such as hydrogen bonding, ionic interactions and Van der 10 Waals interactions.

The chemical-probe approach to ligand design is described, for example, by Goodford, J. Med. Chem. 1985 28 849, the contents of which are hereby incorporated by reference, and is implemented in several commercial software packages, such as GRID (product of Molecular Discovery Ltd., West 15 Way House, Elms Parade, Oxford OX2 9LL, U.K.). Pursuant to this approach, the chemical prerequisites for a site-complementing molecule are identified at the outset, by probing the active site (as represented via the atomic coordinates shown in Fig. 1) with different chemical probes, e.g., water, a methyl group, an amine nitrogen, a carboxyl oxygen, and a hydroxyl. 20 Favored sites for interaction between the active site and each probe are thus determined, and from the resulting three-dimensional pattern of such sites a putative complementary molecule can be generated.

Programs suitable for searching three-dimensional databases to identify molecules bearing a desired pharmacophore include: MACCS-3D and ISIS/3D 25 (Molecular Design Ltd., San Leandro, CA), ChemDBS-3D (Chemical Design Ltd., Oxford, U.K.), and Sybyl/3DB Unity (Tripos Associates, St. Louis, MO).

Programs suitable for pharmacophore selection and design include: DISCO (Abbott Laboratories, Abbott Park, IL), Catalyst (Bio-CAD Corp., Mountain View, CA), and ChemDBS-3D (Chemical Design Ltd., Oxford, 30 U.K.).

Databases of chemical structures are available from a number of sources including Cambridge Crystallographic Data Centre (Cambridge, U.K.) and Chemical Abstracts Service (Columbus, OH).

*De novo* design programs include Ludi (Biosym Technologies Inc., San 35 Diego, CA), Sybyl (Tripos Associates) and Aladdin (Daylight Chemical Information Systems, Irvine, CA).

Those skilled in the art will recognize that the design of a mimetic may require slight structural alteration or adjustment of a chemical structure designed or identified using the methods of the invention.

The invention may be implemented in hardware or software, or a  
5 combination of both. However, preferably, the invention is implemented in computer programs executing on programmable computers each comprising a processor, a data storage system (including volatile and non-volatile memory and/or storage elements), at least one input device, and at least one  
10 output device. Program code is applied to input data to perform the functions described above and generate output information. The output information is applied to one or more output devices, in known fashion. The computer may be, for example, a personal computer, microcomputer, or workstation of conventional design.

Each program is preferably implemented in a high level procedural or  
15 object-oriented programming language to communicate with a computer system. However, the programs can be implemented in assembly or machine language, if desired. In any case, the language may be compiled or interpreted language.

Each such computer program is preferably stored on a storage medium  
20 or device (e.g., ROM or magnetic diskette) readable by a general or special purpose programmable computer, for configuring and operating the computer when the storage media or device is read by the computer to perform the procedures described herein. The inventive system may also be considered to be implemented as a computer-readable storage medium, configured with  
25 a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner to perform the functions described herein.

Compounds designed according to the methods of the present  
invention may be assessed by a number of *in vitro* and *in vivo* assays of  
30 hormone function. For example, the identification of EGF receptor antagonists of may be undertaken using a solid-phase receptor binding assay. Potential antagonists may be screened for their ability to inhibit the binding of europium-labelled EGF receptor ligands to soluble, recombinant EGF  
receptor in a microplate-based format. Europium is a lanthanide fluorophore,  
35 the presence of which can be measured using time-resolved fluorometry. The sensitivity of this assay matches that achieved by radioisotopes,

measurement is rapid and is performed in a microplate format to allow high-sample throughput, and the approach is gaining wide acceptance as the method of choice in the development of screens for receptor agonists/antagonists ( see Apell et.al. J. Biomolec. Screening 3:19-27, 1998 :

5 Inglese et. al. Biochemistry 37:2372-2377, 1998).

Binding affinity and inhibitor potency may be measured for candidate inhibitors using biosensor technology.

The EGF receptor antagonists may be tested for their ability to modulate receptor activity using a cell-based assay incorporating a stably  
10 transfected, EGF-responsive reporter gene (Souriau, C., Fort, P., Roux, P., Hartley, O., Lefranc, M-P., Weill, M., 1997, Nucleic Acids Res. 25:1585-1590). The assay addresses the ability of EGF to activate the reporter gene in the presence of novel ligands. It offers a rapid (results within 6-8 hours of hormone exposure), high-throughput (assay can be conducted in a 96-well  
15 format for automated counting) analysis using an extremely sensitive detection system (chemiluminescence). Once candidate compounds have been identified, their ability to antagonise signal transduction via the EGF-R can be assessed using a number of routine in vitro cellular assays such as inhibition of EGF-mediated cell proliferation. Ultimately, the efficiency of  
20 antagonist as a tumour therapeutic may be tested in vitro in animals bearing tumour isografts and xenografts as described (Rockwell, P., O'Connor, W.J., King, K., Goldstein, N.I., Zhang, L.M., Stein, C.A., 1997, Proc Natl Acad Sci U S A 94:6523-6528; Prewett, M., Rothman, M., Waksal, H., Feldman, M., Bander, N.H., Hicklin, D.J., 1998 Clin Cancer Res 4:2957-2966).

25 Tumour growth inhibition assays may be designed around a nude mouse xenograft model using a range of cell lines. The effects of the receptor antagonists and inhibitors may be tested on the growth of subcutaneous tumours.

#### Comparative modelling

30 The comparative modelling method exploits the observation that proteins with more than 25% amino acid identity will almost always have a similar protein backbone (Sander, C. And Schneider, R., 1991, Proteins: Structure Function and Genetics, 9, 56-68). In some cases, proteins will have similar backbone structures with a lower proportion of identical amino acids.  
35 By aligning the sequence of a (target) protein which is to be modelled with the sequences with known structures (the templates), a model of the protein



can be obtained. Where a region of the target sequence follows the sequences of a template, the backbone of the target is built to follow that of the template. Where the target sequence can not be aligned to a target sequence, the so-called insertion must be constructed by other means (Greer, J., 1991, Meth. Enzym. pp 239-252).

The MODELLER program (Šali, A and Blundell, T.L., 1993, J. Mol. Biol. 234, 779-815) is a semi-automated approach to building models of proteins given the structures of one or more template structures and an alignment between the sequences of the target protein and the templates. Based on the sequence alignment and a set of rules derived from the analysis of sets of aligned structure, the program generates a series of restraints for variables such as C $\alpha$ -C $\alpha$  distances, main chain and side chain dihedral angles for the target structure. The restraints are expressed in terms of probability density functions (PDFs). The PDFs are combined to yield an expression for the most probable structure as a function of the variables (C $\alpha$ -C $\alpha$  distances etc). The program then attempts to find structures to maximise the value of this function. In effect, the program attempts to minimise a transformed version of this function.

While some comparative modelling approaches involve the explicit building of regions of the model for which there is no sequence alignment with a template, the MODELLER program constructs PDFs for these regions, thus including them in the consideration of constructing a comparative model. It is conceivable that once a comparative model has been constructed using MODELLER than an algorithm to build the structures of these regions is applied.

The MODELLER program was used to build the structures of the extracellular portion of the EGF receptor using the 3D structure of the IGF-1 receptor (as described in PCT/AU98/00998) as a template. The description of the generation of these models is outlined below.

### Construction of the alignment

The region of the IGF-1 receptor whose structure is known (Garrett, T.P., McKern, N.M., Lou, M., Frenkel, M.J., Bentley, J.D., Lovrecz, G.O., Elleman, T.C., Cosgrove, L.J., Ward, C.W., 1998 Nature 394:395-399) consists of three domains, the L1 domain, cysteine-rich domain (CRD) and the L2 domain (in order of increasing residue number). The L1 and L2 domains adopt similar folds, each consisting of a single-stranded right-hand  $\beta$ -helix.

The helix contains three  $\beta$ -sheets which make up the left and right sides and the bottom of the  $\beta$ -helix. The top is less regular. This type of  $\beta$ -helix has been dubbed a "breadloaf". The cysteine-rich domain (CRD) consists of eight small modules, each of which has one or two disulfide bonds. The first three  
5 modules of the CRD contain two disulfide bonds which have a Cys1-Cys3 and Cys2-Cys4 disulfide pairing arrangement. The next four have a single disulfide bond with a so-called  $\beta$ -finger structure. The eighth module of the CRD contains one disulfide bond but is not a  $\beta$ -finger.

The sequence of the EGF receptor extracellular domain can be divided  
10 into four domains, L1, S1, L2 and S2 (in order of increasing residue number) on the basis of internal homology and homology with the insulin receptor family (Ward, C.W., Hoyne, P.A., Flegg, R.H., 1995, *Proteins* 22:141-153; Bajaj, M., Waterfield, M.D., Schlessinger, J., Taylor, W.R., Blundell, T., 1987, *Biochim Biophys Acta* 916:220-226). The L1 and L2 domains are similar in  
15 sequence to each other and to the L1 and L2 domains in the IGF-1 receptor. The S1 and S2 domains are similar in sequence and also similar to the CRD of the IGF-1 receptor. These three domains contain a large number of cysteine residues, although the S2 domain of the EGF receptor has two less cysteine residues than does the CRD of the IGF-1 receptor and the S1 domain  
20 of the EGF receptor.

Two important sequence motifs are found in the EGF receptor sequence which are conserved in other EGF receptor homologues. The first motif is the sequence CXXXXXXW which is found near the end of the sequences of the L1 and L2 domains of the EGF receptor and its homologues  
25 where C is cysteine and W is tryptophan. (The motif in the L1 domain of the EGF receptor consists of C133-W140 and in the L2 domain consists of C446-W453.) The second motif is the sequence CW which occurs near the start of the S1 and S2 domains of the EGF receptor (C175-W176 in the S1 domain and C491-W492 in the S2 domain). The two motifs also occur in the insulin  
30 receptor family (C120XXXXXXW127 and C175W176 in IGF-1 receptor) in the L1 domain and cysteine-rich domain respectively. In contrast to the EGF receptor and its homologues the first of these two motifs does not occur in the L2 domain of the insulin receptor family. Structurally, the first motif corresponds to part of the L1 domain which allows penetration of the  
35 tryptophan residue of the second motif into the  $\beta$ -helix. As the first sequence motif is absent from the L2 domain of the IGF-1 receptor, very little of the

structure of this domain was used as a template in the modelling of the EGF receptor.

#### Construction of the alignment of L1 and S1

As the L1 domain of the IGF-1 receptor has a defined core, the  
5 sequence alignment was manually constructed with a view to placing most of  
the conserved hydrophobic residues of the EGF receptor such that their side  
chains point towards the  $\beta$ -helical core. Homologues of the EGF receptor  
were included in the alignment to assist with the identification of such  
residues (Figure 1). Other IGF-1 receptor residues whose positions were  
10 conserved were the four cysteine residues in the L1 domain and the residues  
Arg 77, Trp 127, Trp 176 and Gln 182. Two small regions of the IGF-1  
receptor were also included in the alignment. The first of these regions  
includes the sequence Ser 375 - Lys 380 from the L2 domain of the IGF-1  
receptor and is used as a template for modelling the EGF receptor residues  
15 Asp 51 - Lys 56. Additional flanking residues were also used. Residues Ile  
385 - Phe 397 of the IGF-1 receptor were also used as a template to better  
model the EGF receptor residues Ile 83 - Leu 95 (Figure 1).

The alignment of the S1 domain of the EGF receptor to the cysteine-  
rich domain of the IGF-1 receptor used the same combination of modules. All  
20 of the putative modules of the EGF receptor S1 domain were aligned to part  
or all of the corresponding module of the CRD of the IGF-1 receptor. The  
third module of the IGF-1 receptor CRD (Cys 201 - Cys 218) was used as an  
additional template to the first (Cys 166 - Cys 183) and second (Cys 191 - Cys  
207) putative modules of the EGF receptor S1 domain. The residues Cys 230 -  
25 Cys 246 of IGF-1 receptor, which include the protein's fifth module, were  
aligned to the EGF receptor residues Cys 267-Cys 283 (which include the EGF  
receptor S1 domain's putative sixth module).

#### Construction of the alignment of L2 and S2

Construction of the alignment of the sequence of the L2 domain of the  
30 EGF receptor to the sequence of the L1 domain of the IGF-1 receptor followed  
similar principles to that of the alignment of the L1 domain of the EGF  
receptor. The region Ile 385 - Phe 397 of the IGF-1 receptor served as an  
additional template and its sequence was aligned to Ile 402 - Leu 414 of the  
EGF receptor (Figure 2).

35 An analysis of  $\beta$ -finger modules in the IGF-1 receptor, TNF receptor  
and the laminin- $\gamma$  structures revealed that these modules could be classified

into three types exhibiting some structural and sequence conservation. Two of the structural types are relevant to the IGF-1 receptor and the EGF receptor. The first type of  $\beta$ -finger is characterised by structural conservation of the C-terminal part of the module and also of the linker region after the module. The signature sequence is C...CXXC where the third cysteine residue is the start of another  $\beta$ -finger module. The second type of  $\beta$ -finger is characterised by structural conservation of the N-terminal portion of the module and also of the linker region after the module. The signature sequence is C...CXXXC where the third cysteine is the start of a module whose disulfide bonding pattern has a Cys 1-Cys 3, Cys 2-Cys 4 arrangement.

Comparison of the sequences of the modules of the IGF-1 receptor CRD with the sequence of the EGF receptor S2 domains suggested that the arrangement of modules in the S2 domain were different from those of the IGF receptor CRD and the EGF receptor S1 domain. The residues of the third module in the CRD of the IGF-1 receptor, Cys 201-Cys 218, could be aligned with the segments of the EGF receptor S2 domain sequence: Cys 482 - Cys 499; Cys 534 - Cys 555 and Cys 596 - Cys 612. These modules are the putative first, fourth and seventh modules of the S2 domain. The residues of the first EGF receptor module were also aligned to residues Cys 152 - Cys 181 of the first module of the IGF-1 receptor CRD. The residues of the fourth module in the CRD of the IGF-1 receptor, Cys 221 - Cys 230, a beta-finger module of the first type described above, could be aligned with the regions of sequence Cys 502 - Cys 511 and Cys 558 - Cys 567. These two regions of the EGF receptor S2 domain are the putative second and fifth modules. By elimination, the regions between the two sets of remaining cysteine residues (the putative third and sixth modules) were assigned as  $\beta$ -finger modules of the second type. These regions of sequence are followed by three residues and then a module containing four cysteine residues. The N-terminal regions of the fifth (Cys 234 - Cys 246) and seventh modules (Cys 277 - Cys 291) of the IGF-1 receptor CRD were both aligned to the N-terminal regions of the two modules (Cys 515 - Cys 531 and Cys 571 - Cys 593).

In the IGF-1 receptor CRD, there is no occurrence of a  $\beta$ -finger module being followed by a module containing four cysteine residues. Thus, the positioning of the fourth module in the EGF receptor S2 model relative to the third module is essentially arbitrary. The same applies to the positioning of

the seventh module relative to the sixth module of the EGF receptor S2 domain model.

#### Construction of the model

Version 3 of the MODELLER program (Modeler User Guide, October 5 1996, San Diego Molecular Simulations Inc) was used to build models of the EGF receptor. The various sequences of the IGF-1 receptor and the EGF receptor shown in Figure 1 were used as the alignment for the construction of the model of the L1 and S1 domains of the EGF receptor. The coordinates of each of the IGF-1 receptor sequences (i.e. the templates) shown in Figure 1 10 were also used as input for the MODELLER program. Additional distance restraints were generated between C $\alpha$  atoms of selected residues. The restraints were generated as follows. The small IGF-1 receptor templates were superimposed into the structure of the first two domains of the IGF-1 receptor using the C $\alpha$  atoms of the residues which are aligned in Figure 1. 15 Using the Homology module of the Insight II program (Homology User Guide, October 1995, San Diego BIOSYM/MSI) coordinates were built for the EGF receptor residues which are aligned to the IGF-1 receptor coordinates which are in bold typeface. From these coordinates, distance restraints in the form of Gaussian curves were constructed for pairs of C $\alpha$  atoms with a distance 20 less than 50Å. The sigma value of the Gaussian curves was set to be 2Å. A MODELLER run was submitted using the alignment in Figure 1. The built models of proteins attempt to satisfy these restraints in addition to the restraints the program derives from the alignment.

The aligned IGF-1 receptor and EGF receptor sequences of Figure 2 25 were used as the alignment for creating the model of the L2 and S2 domains of the EGF receptor. The coordinates of the each of the IGF-1 receptor sequences shown in Figure 2 were used as the structural templates. Two separate sets of additional restraints were used. The first set were based on the underlined IGF-1 receptor residues which are aligned to EGF receptor 30 residues Cys 482 – Cys 534 (the first module of the S2 domain to the first cysteine of the fourth module). From the coordinates of the C $\alpha$  atoms of these residues, distance restraints in the form of Gaussian curves were constructed for pairs of C $\alpha$  atoms with a distance less than 50Å. The second set of additional restraints were based on the C $\alpha$  atoms of the underlined IGF-1 35 residues which are aligned to EGF receptor residues Cys 534 – Cys 596 (the fourth module of the S2 domain to the first cysteine of the seventh module).

The signal value of the Gaussian curve used to construct the additional restraints was 1Å.

For both sets of models, the MODELLER program constructed 20 models whose coordinates were perturbed from an initial structure by a random value of maximum distance 4Å. The refinement level used was the 'refine1' option in the MODELLER program.

Most of the insertion regions of the EGF receptor models were constructed using the "loop" routine of version 4 of MODELLER (Modeler User Guide, June 1997, San Diego Molecular Simulations Inc) . Coordinates for each insertion were built using one of the two models obtained in the previous section as a scaffold. The regions of sequence for which coordinates were built in this manner were 1-5, 8-12, 16-23, 46-51, 101-107, 145-148, 184-191, 241-262, 319-328, 522-530, 540-546, 578-600 and 612-621. Coordinates for residues 351-368 and 387-393 were constructed simultaneously due to the proximity of these regions in the model of the L2 domain. For each insertion, 50 models were constructed. In cases where the generated loops with the lowest scores had similar backbone structures, the loop building process was considered to have converged and the coordinates of the loop replaced those of the same residues on the refined model. Where the loop structures did not converge, the structures with the three lowest MODELLER loop scores were evaluated using Procheck (Laskowski RA, MacArthur MW, Moss DS, Thornton JM. (1993). J Appl. Crystallogr 26: 283-291), ProsaII (Hendlich M, Lackner P, Weitckus S, Floeckner H, Froschauer R, Gottsbacher K, Casari G, Sippl MJ. (1990) J Mol Biol 216:167-180.; Sippl MJ. (1993) Proteins 17: 355-362.) and Profiles-3D (Bowie JU, Lüthy R, Eisenberg D. (1991) Science 253:164-170.; Lüthy R, Bowie JU, Eisenberg D. 1992. Nature 356:83-85.). For several of these loops, the one with the second lowest MODELLER score was selected as it had a more favorable Profiles3D and ProsaII plot.

In order to retain certain secondary structures, additional restraints were used in the construction of some of the loops. Restraints with the form of a right-handed half-Gaussian function with a  $s$  value of 0.05Å were used to hold selected mainchain N-O distances to 3.0Å or less. The atom pairs for which this additional restraint was added were: Gln 139.N - Gln 184.OE1, Val 268.N - Tyr 261.O, Val 268.O - Tyr 261.N, Ser 506.N - Ser 529, Ile 562.N - His 591.O and Glu 578.N - Val 592.

### Structure of the EGF receptor model

The structure of the L1 and S1 domains of the EGF receptor as determined by the modelling described above is shown in Figure 3, while the structure of the L2 and S2 domains is shown in Figure 4. The superposition of these two models onto the structure of the extracellular domains of the IGF-1 receptor is shown in Figure 5.

The coordinates of the EGF receptor domains L1, S1, L2 and S2 are shown in Figure 6.

Figures 7, 8 and 9 show the sidechains of residues of the EGF receptor models which face the large cavity as shown in Figure 5. Figure 10 shows the sidechains of residues of the face of the EGF receptor L2 domain which contains the second beta sheet (the lower face of the L2 domain using the orientation shown in Figure 4).

The structures of the L1 and S1 domains are similar to those of the IGF-1 receptor structure, as expected. There are three major differences in the S1 domain of the EGF receptor model from the structure of the IGF-1 receptor cysteine-rich domain. The first module of the S1 domain is noticeably smaller than that of the IGF-1 receptor CRD. The sixth module (Cys 271 – Cys 283) of the S1 domain is smaller than that of the IGF-1 receptor and occupies less of the region between the L1 and L2 domains. The fifth module (Cys 240 – Cys 267) contains a large insertion which points away from the L1 domain. The eighth module of the EGF receptor S1 domain (Cys 305 – Cys 309) and the linker region (Arg 310 – Val 312) which follows it are similar in structure to the analogous regions of the IGF-1 receptor. Like the IGF-1 receptor, the linker region is postulated to be a hinge region about which the S1 domain and the L2 domain can reorient.

A region of the EGF receptor in the L2 domain which could not be aligned with the IGF-1 receptor includes the residues Trp 386 – Pro 387 which are conserved across the EGF receptor family. This sequence motif is not found in the insulin receptor family and may represent a region of novel structure.

The amino acids 352-367 correspond to a large insertion in the L2 domain of the EGF receptor. The amino acids 351-364 have been identified as the epitope for several antibodies against the EGF receptor (Wu, D.G., Wang, L.H., Sato, G.H., West, K.A., Harris, W.R., Crabb, J.W., Sato, J.D., 1989, J. Biol. Chem. 264:17469-17475). This region forms a loop which sticks out of the

surface is consistent with this region being accessible to antibodies. The structure itself is difficult to model accurately since its sequence does not correspond to any part of the IGF-1 receptor sequence. The position of this insertion is in approximately the same region as where the IGF 1 receptor differs in backbone structure.

The S2 domain model of the EGF receptor adopts a different arrangement of modules and consequently a different shape that of the CRD of the IGF-1 receptor and the S1 domain model of the EGF receptor. The disulfide bond arrangement is the same as that predicted by similarity to the tumour necrosis receptor (Ward, C.W., Hoyne, P.A., Flegg, R.H., 1995, Proteins 22:141-153) and has since been confirmed by mass spectroscopic analyses of proteolytically digested EGF receptor extracellular domain (Abe, Y., Odaka, M., Inagaki, F., Lax, I., Schlessinger, J., Kohda, D., 1998, J. Biol. Chem. 273:11150-11157). The only significant contact of the S2 domain with the L2 domain of the EGF receptor model is the intercalation of Trp 492 into the L2 domain, analogous to that made by Trp 176 in the S1 domain of the EGF receptor and Trp 176 in the CRD of the IGF-1 receptor to their respective L1 domains. Unlike the S1 domain of the EGF receptor, the rest of the S2 domain does not make any contacts with the L2 domain. The S2 domain is rod-like and points out from the L2 domain with a different geometry to the manner in which the S1 domain points out from the L1 domain.

#### Putative binding sites of the EGF receptor

From the IGF-1 receptor structure and a number of insulin receptor mutants, one of the regions of insulin binding was proposed to the face of the L1 domain which contains the second  $\beta$ -sheet (Garrett, T.P., McKern, N.M., Lou, M., Frenkel, M.J., Bentley, J.D., Lovrecz, G.O., Elleman, T.C., Cosgrove, L.J., Ward, C.W., 1998 Nature 394:395-399). This surface is characterised by a number of hydrophobic residues which point out of the structure and also the presence of a structurally conserved loop. By analogy, we propose that the analogous  $\beta$  sheets of the L1 and L2 are potential binding sites. These sheets contain a number of hydrophobic residues, conserved amongst EGF receptor family members, which point away from the core of the  $\beta$ -helix structure. Residue 45 of a mutant EGF has been cross-linked to the residue Lysine 465 which is in the last strand of the lower  $\beta$  sheet of the L2 domain. (Summerfield, AE et al, J Biol Chem, 1996, 271(33), 19656-19659). Tyrosine 101 has been cross-linked to the N-terminus of EGF (Woltjer, RL et al, PNAS,



1992, 89(16), 7801-7805). This residue is in the portion of sequence which immediately follows a strand in the lower  $\beta$  sheet of L1.

The side chain of asparagine 1 of EGF has been cross-linked to lysine 336 of the EGF receptor (Wu, DG et al, PNAS, 1990, 87(8), 3151-3155). The latter residue is in the N-terminal helix of the L2 domain and points towards the cavity which is formed when the two halves of the EGF receptor are positioned in a similar arrangement to the first three domains of the IGF-1 receptor. Two nearby residues, Asn 328 and Asn 337 are glycosylated. This mutation is in a similar position to the insulin receptor mutant S323L which has aberrant insulin binding.

Several insertional mutants of the EGF receptor extracellular domain were constructed to probe the role of several regions of the receptor (Harte, M.T., Gentry, L.E., 1995, Arch Biochem Biophys 322:378-389). A number of these mutants were not detectably secreted by the cells producing them, suggesting that they did not fold to form stable proteins. Most of these insertions were in positions in the model structure where they would be unable to tolerate an insertion. In contrast, most of the other insertions were in loops or other positions which, according to the model, are able to tolerate insertions. EGF receptor extracellular domain mutants with insertions at residues 162, 169, 174 and 220 bound EGF with a similar affinity to the wild-type EGF receptor extracellular domain but bound TGF- $\alpha$  with a lower affinity. The first of these insertions was located one residue before the last cysteine residue of the L1 domain. The second and third insertions were present in the first module of the EGF receptor S1 domain and the fourth was present in the third module of the S1 domain. All of these positions are on a side of the molecule far removed from the large cavity as shown in Figure 5. EGF receptor mutants with insertions at positions 251 (in the fifth module of the S1 domain) and 575 (in the sixth module of the S2 domain) appeared to bind twice as much ligand as the wild-type receptor. Two insertional mutants which showed reduced EGF receptor binding contained insertions at positions 291 (in the seventh module of the S1 domain) and 474 (one residue before the last cysteine of the L2 domain).

Another EGF receptor mutant which shows altered ligand binding behaviour is the R497K mutant. The site of this mutation in the first module of the S2 domain and faces the side of the L2 domain opposite to that containing residue 465. This mutant binds EGF in a similar fashion as wild-

type receptor but abolishes the high affinity binding site for TGF- $\alpha$  (Moriai, T., Kobrin, M.S., Hope, C., Speck, L., Korc, M., 1994, Proc Natl Acad Sci U S A 91:10217-10221).

On the faces containing the second  $\beta$ -sheet (the lower face according to the orientations shown in Figures 3 and 4) of the L1 and L2 domains are a number of solvent-exposed hydrophobic residues including Tyr 64, Leu 66, Tyr 89, Tyr 93, Leu 348, Phe 380 and Phe 412. According to a survey of protein-protein interfaces, tyrosine, phenylalanine and leucine are more likely to be involved in an interface than on the exterior of a protein complex (Tsai C-J, Lin SL, Wolfson, HJ, Nussinov R (1997) Protein Sci 6: 53-64). Lys 465 is located on the lower face of the L2 domain and Tyr 101 is proximal to the lower face of the L1 domain and are consistent with the lower faces of the domains having roles in ligand binding.

#### Strategies for developing EGF receptor ligands

For several signalling systems, ligand analogues which have antagonist properties have been described. These ligands include the human growth hormone (Chen WY, Chen NY, Yun J, Wagner TE, Kopchick JJ (1994) J Biol Chem 269:15892-15897), interleukin-6 (Savino R, Lahm A, Salvati AL, Ciapponi L, Sporeno E, Altamura S, Paonessa G, Toniatti C, Ciliberto G EMBO J 1994 Mar 15;13(6):1357-67) and interleukin-4 (Kruse N, Tony HP, Sebald W (1992) EMBO J 11:3237-3244; Zurawski SM, Vega F Jr, Huyghe B, Zurawski G (1993) EMBO J 12:2663-2670). The function of these unmodified ligands is to bind their receptors and then subsequently recruit a second receptor molecule. The mutations of the ligands mentioned above are in positions which interfere with the binding of the second receptor (de Vos AM, Ultsch M, Kossiakoff AA (1992) Science 255:306-312; Brakenhoff JP, de Hon FD, Fontaine V, ten Boekel E, Schooltink H, Rose-John S, Heinrich PC, Content J, Aarden LA (1994) J Biol Chem 269:86-93; Davis ID, Treutlein HR, Friedrich K, Burgess AW (1995) Growth Factors 12:69-83).

To date, no analogues of EGF receptor ligands have been found which are purely antagonistic. Whether EGF and its homologues have sites of binding for two receptor molecules, like the proteins described above, has not been shown. Analysis of <sup>1</sup>H NMR transferred nuclear Overhauser enhancement data for titration of TGF- $\alpha$  with the extracellular domain of the EGF receptor indicates that most parts of the ligand are in contact with the receptor upon binding (McInnes C, Hoyt DW, Harkins RN, Pagila RN,

Debanne MT, O'Connor-McCourt M, Sykes BD (1996) J Biol Chem 271:32204-32211). However, the concentrations used in the experiment were such that the dominant receptor species was the ligand-receptor complex with 2:2 stiochiometry. However, even if the ligands of the EGF receptor are buried in the cleft formed by the first three domains of the receptor, it is difficult to envisage that such binding will lead to contact with most of the bound ligand when only one receptor binds the ligand. In an alternative scheme, at least two separate faces on EGF are required to bind into the large cleft of a single EGF receptor molecule which enacts a conformational change in the receptor which then allows it to dimerise. An antagonist may bind to the first binding site of the receptor and not the second, thus preventing dimerisation and subsequent signalling of the receptor. Thus, delineation of the parts of the ligand involved in the (putative) primary and secondary binding faces would greatly assist antagonist design.

Using the EGF receptor model and the known structures of EGF receptor ligands, it may be possible to construct a model, or a partial model, of ligand binding which could suggest which parts of bound ligand are involved in binding to the first and second EGF receptors of the ligand-receptor complex. There are several computer programs that can assist with the construction of such models. Programs such as Quilt (Lijnzaad P, Argos P (1997) Proteins 28:333-343 ; Lijnzaad P, Berendsen HJ, Argos P (1996) Proteins 26:192-203; Lijnzaad P, Berendsen HJ, Argos P 1996 Proteins 25:389-397) can be used to suggest sites on proteins involved in interactions with other proteins. Possible structures of protein complexes can be obtained by programs such as FT-DOCK (Gabb HA, Jackson RM, Sternberg MJ (1997) J Mol Biol 272:106-120) and GRAMM (Vakser IA (1996) Biopolymers 39:455-464; Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA (1992) Proc Natl Acad Sci U S A 89:2195-2199). The calculation of electrostatic potentials from the Poisson-Boltzmann equation has been used to investigate complexes made up of cytokines and growth factors and their receptors (Demchuk E, Mueller T, Oschkinat H, Sebald W, Wade RC (1994) Protein Sci 3:920-935) and may guide the construction of model complexes. The construction of models will suggest regions of the EGF receptor ligands which may be involved in receptor binding. With the model and supporting experiments, it is envisaged that mutants of EGF and TGF- $\alpha$  will be constructed which are potential antagonists.

The majority of targets for drugs which have made use of structural information are enzymes. One advantage of enzymes over other types of proteins is the presence of substrate-binding clefts whose normal function is to bind small molecule substrates or short lengths of peptides. In contrast, few small molecule inhibitors have been developed which inhibit protein-protein interactions.

Desolvation of protein surfaces appears to be an important factor in the formation of a protein-protein complex. Since, unlike the substrate-binding clefts of enzymes, protein-binding surfaces tend to be much less concave, a bound small molecule is unlikely to provide enough desolvation to enable tight binding. The lower surfaces of the L1 and L2 domains, which have been suggested to be involved in ligand binding, contain hydrophobic regions which suggest that they need to be buried for strong binding of a molecule to these surfaces to occur. We envisage that cyclic molecules, including cyclic peptides, may be able to bind to such surfaces. Hydrophobic functional groups may be chosen which, when bound to the hydrophobic regions of the relevant face, desolvate regions of the protein. Some of the functional groups which interact with the protein will be polar or charged to make favourable electrostatic interactions. Other parts in the cyclic molecule may be polar or charged to increase the aqueous solubility of the molecule. Cyclic molecules also have the advantages of having few possible conformations when unbound, providing a lower loss of entropy upon binding and thus greater binding as compared to a non-cyclic analogue. A degree of flexibility would exist and would allow the molecule to alter its conformation to better accommodate the protein it is binding to.

Algorithms such as LUDI (Bohm HJ (1992) J Comput Aided Mol Des, 6: 593-606) can be used to search for functional groups and molecular moieties which may interact with a surface of the EGF receptor model. Algorithms such as CLIX (Lawrence MC, Davis PC (1992) Proteins 12:31-41) or DOCK (Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) J Mol Biol 161:269-88) can be used to search a database of molecular structures for those which have shape and/or chemical complementarity to the EGF receptor. Computational combinatorial design algorithms (Miranker A, Karplus M Proteins (1991) 11:29-34; Eisen MB, Wiley DC, Karplus M, Hubbard RE; Caffisch A (1994) Proteins 19:199-221) can also be tried. In one instance, a combinatorial approach has been used to design peptides to

inhibit the interaction of the proteins Ras and Raf (Zeng, J. et al, Protein Engineering, to be published).

We envisage that as an alternative to a cyclic molecule, a small protein could be used as a scaffold for placing amino acids that will interact with the EGF receptor. At least one small protein (potato carboxypeptidase inhibitor) with a fold different to that of EGF receptor ligands has been identified which is a weak EGF antagonist (Blanco-Aparicio C, Molina MA, Fernandez-Salas E, Frazier ML, Mas JM, Querol E, Aviles FX, de Llorens R (1998) J Biol Chem 273:12370-12377). The use of a structural scaffold for proteins with diverse functions has been observed in Nature (Lin SL, Nussinov R 1995 Nat Struct Biol 2:835-837). Other molecular scaffolds such as dendrimers may also be considered which can be used to present the functional groups which will tightly interact with the EGF receptor.

At least two, non-exclusive modes of action can be envisaged. The first mode involves a molecule competing for binding sites with one of the EGF receptor's natural ligands. Most likely, the molecule will prevent the receptor dimerisation which is required for activation of the EGF receptor, thus acting as an antagonist. We do not rule out the possibility that the binding may be activating and the molecule acts as an agonist. The second potential mode of action is for the molecule to bind to a site on the EGF receptor which is not necessarily a ligand binding site. Such a molecule may be physically large enough to hinder physical access of a second receptor to the receptor which binds the molecule in question. This would hinder dimerisation and subsequent activation of the receptor. If the molecule is sufficiently "sticky", it may attract a second EGF receptor and induce dimerisation, thereby acting as an agonist rather than an antagonist.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.